

APPLICATION FOR UNITED STATES LETTERS PATENT

For

STORAGE RESOURCE MANAGEMENT ACROSS MULTIPLE PATHS

Inventors:

Vijay Deshmukh

Benjamin Swartzlander

Timothy J. Thompson

Prepared by:

BLAKELY SOKOLOFF TAYLOR & ZAFMAN LLP

32400 Wilshire Boulevard

Los Angeles, CA 90025-1026

(408) 720-8300

Attorney's Docket No.: 005693.P052

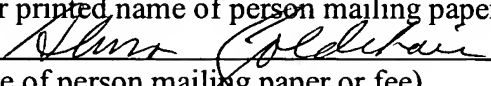
"Express Mail" mailing label number: EV 34163935 US

Date of Deposit: March 12, 2004

I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria VA 22313-1450

Alma Goldchain

(Typed or printed name of person mailing paper or fee)


(Signature of person mailing paper or fee)

3/12/04
(Date signed)

STORAGE RESOURCE MANAGEMENT ACROSS MULTIPLE PATHS

FIELD OF THE INVENTION

[0001] At least one embodiment of the present invention pertains to networked storage systems, and more particularly to a method and apparatus for collecting and reporting data pertaining to files stored on a storage server.

BACKGROUND

[0002] A file server is a type of storage server which operates on behalf of one or more clients to store and manage shared files in a set of mass storage devices, such as magnetic or optical storage based disks. The mass storage devices are typically organized as one or more groups of Redundant Array of Independent (or Inexpensive) Disks (RAID). One configuration in which file servers can be used is a network attached storage (NAS) configuration. In a NAS configuration, a file server can be implemented in the form of an appliance, called a filer, that attaches to a network, such as a local area network (LAN) or a corporate intranet. An example of such an appliance is any of the NetApp Filer products made by Network Appliance, Inc. in Sunnyvale, California.

[0003] A filer may be connected to a network, and may serve as a storage device for several users, or clients, of the network. For example, the filer may store user directories and files for a corporate or other network, such as a LAN or a wide area network (WAN). Users of the network can be assigned an individual directory in which they can store personal files. A user's directory can then be accessed from computers connected to the network.

[0004] A system administrator can maintain the filer, ensuring that the filer continues to have adequate free space, that certain users are not monopolizing storage on the filer, etc. A Multi-Appliance Management Application (MMA) can be used to monitor the storage on the filer. An example of such an MMA is the Data Fabric Monitor (DFM) products made by Network Appliance, Inc. in Sunnyvale, California. The MMA may provide a Graphical User Interface (GUI) that allows the administrator to more easily observe the condition of the filer.

[0005] The MMA needs to collect information about files stored on the filer to report back to the administrator. This typically involves a scan or “file walk” of storage on the filer. During the file walk, the MMA can determine characteristics of files stored on the filer, as well as a basic structure, or directory tree, of the directories stored thereon. These results can be accumulated, sorted, and stored in a database, where the administrator can later access them. The MMA may also summarize the results of the file walk so they are more easily readable and understood by the administrator.

[0006] A filer may manage a volume storing several million files. The amount of time and system resources needed to scan such a large volume can make it prohibitive for the MMA to perform the file walk. However, the system administrator still needs the file walk information to effectively manage the filer. What is needed is a way to more effectively monitor a filer that manages a large number of files.

SUMMARY OF THE INVENTION

[0007] Other aspects of the invention will be apparent from the accompanying figures and from the detailed description which follows.

[0008] A method for performing a file walk of a storage server is disclosed. A first path and a second path on a storage server are determined. A first information about the first path is collected using a first agent, and a second information about the second path is collected using a second agent. The first and second information are stored in a common format.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] One or more embodiments of the present invention are illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

[0010] **Figure 1** illustrates a monitoring system for a storage server;

[0011] **Figure 2** illustrates a block diagram of an agent;

[0012] **Figure 3A** illustrates a directory structure represented by a tree;

[0013] **Figure 3B** illustrates the names of the directories shown in the tree 300;

[0014] **Figure 3C** illustrates a tree divided into two paths;

[0015] **Figure 4** illustrates a system for a file walk using several agents and several filers;

[0016] **Figure 5** is a flow chart illustrating a process for performing a file walk across multiple paths;

[0017] **Figure 6A** illustrates a table showing information across a first path;

[0018] **Figure 6B** illustrates a table showing information across a second path;

[0019] **Figure 7A** illustrates a table showing cumulative information for a first path;

[0020] **Figure 7B** illustrates a table showing cumulative information for a second path; and

[0021] **Figure 7C** illustrates a table showing cumulative information for both a first and a second path.

DETAILED DESCRIPTION

[0022] Described herein are methods and apparatuses for storage resource management across multiple paths. Note that in this description, references to “one embodiment” or “an embodiment” mean that the feature being referred to is included in at least one embodiment of the present invention. Further, separate references to “one embodiment” or “an embodiment” in this description do not necessarily refer to the same embodiment; however, such embodiments are also not mutually exclusive unless so stated, and except as will be readily apparent to those skilled in the art from the description. For example, a feature, structure, act, etc. described in one embodiment may also be included in other embodiments. Thus, the present invention can include a variety of combinations and/or integrations of the embodiments described herein.

[0023] According to an embodiment of the invention, one or more filers is managed by a multi appliance management application (MMA). The MMA controls one or more agents which perform a file walk of the filers. The MMA may divide the directory structure of a filer into multiple paths, so that more than one agent can perform file walk of a single filer. A filer may also be scanned by one or more agents having different file systems. For example, a single filer may scanned by the first agent using a first file system, such as the Common Internet File System (CIFS), and a second agent using a separate file system, such as the Network File System (NFS). A directory structure may be represented using a logical tree. The MMA can divide the tree into one or more sub trees. Each of these sub trees can be scanned by a different agent. Each of these sub trees may represent a path. The MMA can divide the directories on a filer into several different paths, so that several different agents may scan a single filer in order to reduce

the amount of time required to complete a file walk. As a result, multiple paths are used to improve storage resource management.

[0024] The MMA is generally a single server that is used to allow a system administrator to monitor a storage or file server. When a high capacity storage server is monitored, the MMA may have difficulty performing its monitoring duties and a file walk at the same time. In fact, the file walk may make the MMA inaccessible to the system administrator, and the MMA may further become a bottleneck to the file walk process, since it may be incapable of performing the file walk in a reasonable amount of time. According to an embodiment of the invention, independent agents are used to perform the file walk, to reduce the load on the MMA. At a later time, the system administrator may want summarized information about the file server. Instead of having the MMA summarize the information, the summaries are compiled by the agent during the file walk, and stored on a database server.

[0025] **Figure 1** illustrates a monitoring system for a storage server. The system 100 includes a filer 102, an MMA 104 including a monitor 106, a database 108, a graphical user interface (GUI) 110, and two agents 112 and 114. The agents 112 and 114 can perform a file walk of the filer 102 for the MMA 104. An agent may be an independent server that is attached to the network and is dedicated to performing file walks. By having an agent perform this task rather than having the MMA do it, the MMA can save its resources for other tasks, such as monitoring current activity on the filer 102 using the monitor 106. Ultimately, one goal is to minimize the amount of work the MMA is required to do. Additionally, multiple agents can be added to perform a complete file walk in less time.

[0026] According to one embodiment of the invention, the agents 112 and 114 may use a file system different from the one used by the filer 102. For example, the agent 112 uses the Common Internet File System (CIFS), while the agent 114 uses the Network File System (NFS). Here, either agent 112 or 114 is able to perform the file walk of the filer 102, regardless of the file system used by the filer 102. The agent 112 also has storage 116 to store the results of a file walk while the walk is occurring and before they are transferred to the MMA 104. The agent 114 may also have attached storage for this purpose.

[0027] The filer 102 is generally attached to a volume 118. The volume 118 may include one or more physical hard drives or removable storage drives that comprise the storage for the filer 102. For example, the volume 118 may comprise a RAID structure. The filer 102 may also be connected to other volumes that comprise storage. A file walk generally scans all files stored on the entire volume 118, regardless of whether all of the files are stored on the same physical drive. Further, although the volume 118 may contain several separate physical drives, the volume 118 may appear and function as a single entity.

[0028] The results of a file walk may be transferred to and stored on the database server 108 after the file walk is complete. The database server 108 can then be accessed by the GUI 110, so that an administrator can search the results of the file walk. The GUI 110 may allow the administrator to easily parse the results of a specific file walk, including allowing the administrator to monitor the total size of files stored on the filer, the size of particular directories and their subdirectories, the parents of specific directories, etc. These queries will be discussed in more detail below. The file walk may

also collect statistics about the files on the filer, such as the total size of files, the most accessed files, the types of files being stored, etc. According to one embodiment, the GUI 110 may be a web-based Java application.

[0029] According to an embodiment of the invention, the summary is written to the database server 108 as a table or a histogram. The summary may then be accessed through a Java applet using a web browser such as Internet Explorer or Netscape. In another embodiment, the summaries are accessed using other programs. Although tables and histograms are shown here, it is understood that any appropriate manner of presenting the summary data to the administrator may be used.

[0030] **Figure 2** illustrates a block diagram of an agent. The agent 112 includes a processor 202, a memory 204, a network adapter 206, and a storage adapter 208. These components are linked through a bus 210. The agent 112, as shown in **Figure 2**, is typical of a network server or appliance, and it is understood that various different configurations may be used in its place. The agent 114 may be similar.

[0031] The processor 202 may be any appropriate microprocessor or central processing unit (CPU), such as those manufactured by Intel or Motorola. The memory 204 may include a main random access memory (RAM), as well as other memories including read only memories (ROM), flash memories, etc. The operating system 212 is stored in the memory 212 while the agent 112 is operating. The operating system includes the file system, and may be any operating system, such as a Unix or Windows based system. The network adapter 206 allows the agent 112 to communicate with remote computers over the network 214. Here, the agent 112 will be collecting data from

the filer 102 and sending data to the MMA 104. The storage adapter 208 allows the agent 112 to communicate with the storage 116 and other external storage.

[0032] Several agents 112 and 114 may be added in order to reduce the amount of time required to file walk a filer 102. The administrator, using the GUI 110 can configure the number of agents 112 and 114 assigned to a file walk. For example, in one embodiment, a single agent 112 or 114 may be able to scan five million files per hour. If the filer 102 has five million files, a single agent 112 or 114 can complete a full file walk of a filer 102 in one hour. However, the administrator may need the file walk information less time. If both agents 112 and 114 are assigned to walk the filer 102, the results of the file walk could be reported within approximately thirty minutes.

[0033] In a further embodiment, the GUI 110 may include an option such that an administrator can specify the amount of time in which the walk should be completed. For example, an administrator may specify that a file walk should be completed in one hour. The MMA 102 can then determine the number of agents 112 or 114 required to perform the file walk within approximately that time period based on the speed of the agent(s) and the number of files on the filer 102.

[0034] **Figure 3A** illustrates a directory structure represented by a tree. The tree 300 includes several nodes 301 through 310. Each node 301 through 310 may represent an individual directory stored on a filer 102. The tree 300 provides a convenient visual representation of directories stored on the filer 102. **Figure 3B** illustrates the directory names of the nodes 301-110 shown in the tree 300. As can be seen, each directory that is stored within another directory is located beneath that directory in the tree. For example, the node 301 represents the directory /u/. The node 302 represents the directory

/u/employees/. The node 301 is considered the parent of its child node 302. In fact, in the tree 300, the node 301 is the parent of all the other nodes 302 through 310. The node 301 may also be referenced to as the “root node.” Nodes that are on the same level of the tree 300, such as the nodes 302 and 307, are known as “siblings” since they both have the same immediate parent.

[0035] The nodes 301-310 are also assigned identification (ID) numbers. The ID numbers are assigned to the nodes 301 through 310 in order starting with the number 1. The ID numbers can be used to easily identify specific directories. The ID numbers as shown in the tree 300 are in a Depth First Search (DFS) order. However, it is understood that other numbering conventions may also be used. The DFS order assigns an ID number to a node by traversing the tree to the bottom first and then across the tree. In this way, all the children of a specific node are assigned ID numbers before a sibling of that node is assigned its ID number. If a node has no more siblings, the process moves back up the tree. The ID numbers are assigned during a file walk, in consecutive order. So, the file walk is also conducted in a DFS order. The DFS order facilitates efficient queries about specific directories and their relationships with other directories.

[0036] A path is a portion of a directory structure on a storage device. For example, a path may be a directory and all of its subdirectories, and here will be a sub tree. **Figure 3C** illustrates a tree divided into two paths. The tree 370 shows the same directory structure represented in the tree 300, however the tree 370 has been divided into two paths 372 and 374. Each path 372 and 374 is a sub tree of the larger tree 370. An administrator can determine the paths 372 and 374 based on the approximate number of files located within a specific subdirectory. For example, the new root node 302 of the

path 372 is the directory /u/employees/. The new root node of the path 374 is the directory /u/administrators/. An administrator may determine that these two sub trees have a roughly equal number of files located within them. Therefore, each path 372 and 374 can be independently scanned by a separate agent 112 or 114. This way, the administrator can reduce the amount of time required to walk the entire tree by having multiple agents 112 and 114 perform the file walk. An agent 112 or 114 treats each path 372 and 374 as if it were an independent filer during the file walk.

[0037] Since the ID numbers of the nodes 301 through 310 are assigned during a file walk, a nodes in the path 372 may have an ID number that is the same as the ID number of a node the from the path 374. Each path or sub tree 372 and 374 will have an independent set of ID numbers. For example, the nodes 302 and 307 may both be assigned same ID number 1, since they are both the root nodes for the paths 372 and 374, respectively. Since the ID numbers can be used to perform queries across the tree 370, the administrator may choose a numbering convention that identifies the nodes individually. For example, the administrator may assign all the nodes in the left path 372 ID numbers beginning with a leading 'L', and may assign all the ID number in the right path 374 with a leading 'R'. This way, the ID numbers can be assigned without confusion as to which node or directory they refer.

[0038] **Figure 4** illustrates a system for a file walk using several agents and several filers. The MMA 102 controls several agents 112, 114, and 404. The agents 112 and 404 use the CIFS file system, and the agent 114 uses the NFS file system. The filers 104 and 402 are by all system independent appliances. In other words, the agents 112, 114 and 404 may scan the filers 104 and 402 independent what file system they are using.

Further, the information and summaries are stored in a neutral format that can easily be read by either file system. The format includes simple information about the files that is represented in the tables shown below. The results of a file walk may be reported in a table or histogram format which is independent of the file system with the agent. This is because of the type of data that is reported back by the agents 112, 114 and 402. The data reported back by the agents typically includes size of files, name of files, location of files, etc. The size of the files may be represented by an integer or other number that is easily stored in a table and that may be easily portable between file systems. Therefore, since the data being reported back by the agents 112, 114 and 404 is independent of the type of file system being used on the filers 402 and 102, the agents 112, 114 and 404 are able to walk any filer.

[0039] **Figure 5** is a flow chart illustrating a process for performing a file walk across multiple paths. The process 500 explains how a file system may be divided into several paths and how data may be collected regarding those files. In block 502, a first path on the storage server is determined. The first path may be determined by an administrator or an MMA 104. The first path may be the path 372 illustrated in **Figure 3C**. The first path may be chosen so that it includes roughly the same number of files as the second path. The number of files located within the first path is more relevant to the amount of time required for walking those files, since each file must be scanned independently, and data retrieved from that analysis will typically be independent of file size.

[0040] In block 504, a second path on the storage server is determined. The second path may comprise the remainder of the file system. The second path may also contain roughly equal the number of files that the first path has. In another embodiment, the

MMA 104 or an administrator may determine a third or further path for the file walk.

This determination may be made depending on the number of agents available for the file walk, the size of the storage server and the number of files stored on the storage server, and the amount of time in which the administrator wishes to complete file walk.

[0041] In block 506, the first path is scanned using a first agent, thereby collecting a first information about the first path. The scan will reveal details about files stored in the first path, such as the file names, size of the files, location of files, the location of directories relative to each other, etc. The first agent will be dedicated to the first path, and will ignore the second path. The information collected may be stored in a table or histogram, which can later can be reviewed by an administrator or an MMA to make decisions about the operation of the filer 102. In block 508, the second path is scanned using a second agent, thereby collecting a second information about the second path. This process is described in block 506, and the information collected by the second agent is similar in scope to the information collected by the first agent. The second agent, likewise, scans the second path while disregarding the first path. In this way, a filer 102 may be scanned in an expedited manner.

[0042] In block 510, the first and second information are stored using a common format. By storing the information from both paths in common format, the MMA 104 and administrator may easily parse that information at a later time. The information may also be linked together, so that an administrator and the MMA 104 may gain an insight as to the overall state of the filer 102. The two agents may be scanning the two paths for the same basic information, which may be reported as two tables or two histograms. Since the formats will be the same, the tables may either be merged at a later time, or linked

together so that total system information may be determined. For example, if a administrator wanted to determine what the large file is stored in the two different paths was, the administrator can instruct the MMA 104 to examine the two different tables compiled by the two different agent. The MMA 104 could compare the large file found on the first path with the large file found on the second path, and easily determine the large file on the entire filer 102.

[0043] **Figures 6A and 6B** illustrate tables of collected information. **Figure 6A** illustrates a table showing information across a first path. The table 600 includes information collected while scanning the first path 372. The table has several columns, including a column 602 listing the directory name, a column 604 listing the number of files in the directory, a column 606 listing a total size of files in the directory, a column 608 listing an average time of the last access to files in the directory, and a column 610 listing the ID number of the directory. **Figure 6B** illustrates a table showing information across a second path. The information contained in the table 600 is generated during the file walk. Likewise, the information contained in the table 650 is generated during the file walk of the second path 374. The table 650 also includes several columns 652-660, which are similar to the columns of the table 600. As mentioned above, the ID numbers found in the columns 610 and 660 include a leading 'L' or 'R' to signify which path the directory belongs to.

[0044] The tables 600 and 650 include information that is specific to the directories listed therein. The tables 650 can easily be appended to the table 600 to create a single table for the filer 102, since each directory has its own information. However, the file walk may also generate universal tables, listing such data as the largest file found, the

oldest file found, etc. **Figures 7A-C** illustrate tables including cumulative information about files on a filer 102. **Figure 7A** illustrates a table showing cumulative information for a first path. **Figure 7B** illustrates a table showing cumulative information for a second path. **Figure 7C** illustrates a table showing cumulative information for both a first and a second path. The table 700 includes interesting files found in the first path 372, and the table 720 includes interesting files found in the second path 374. The rows 702, 704, 706, and 708 list, respectively, the least recently accessed, largest, smallest, and oldest files found in the first path 372. Likewise, the rows 722-728 list the corresponding files found in the second path 374. The table 740 shows the interesting files for the entire filer 102. The rows 742-748 are analogous to the rows 702-708. For example, since the largest file is found in the first path 372, that is the largest file listed in the table 740. On the other hand, the smallest file is found in the second path 374, and that is the file listed in the table 740. It is understood that the tables 700, 720, and 740 may also present a list of the 'n' largest, smallest, oldest, and least recently accessed files. It is also understood that other categories may be used in accordance with the administrator's wishes.

[0045] The table 740 is a combined list of interesting files covering both paths 372 and 374. According to an embodiment of the invention, the MMA 104 may present the table 740 to the administrator, since the administrator may only want to know the interesting files for the entire filer 102, rather than the interesting files for each individual path 372 or 374. The agents 112 and 114 can still save the tables 700 and 720 to the database server 108 after the file walk, and a combined table 740 can be created either later or at the same time. The combined table 740 can either be created on the fly, when the administrator requests it, or can be created following the file walk of the two paths

372 and 374. Since the combined table 740 typically includes relatively few listed files, a relatively small amount of resources is required to form the combined table 740.

[0046] It is understood that other forms of representing the data collected during the file walk may be used. For example, a histogram may represent the usage of several different users. Likewise, a histogram or table may be created that shows the percentage of storage space on a filer 102 occupied by certain types of files. The MMA 104 may be configured so that useful data of any kind can be collected by the agents 112 and 114 and relayed to the administrator. The GUI 110 may also include an interface to allow the administrator to create customized tables or histograms.

[0047] The techniques introduced above have been described in the context of a NAS environment. However, these techniques can also be applied in various other contexts. For example, the techniques introduced above can be applied in a storage area network (SAN) environment. A SAN is a highly efficient network of interconnected, shared storage devices. One difference between NAS and SAN is that in a SAN, the storage server (which may be an appliance) provides a remote host with block-level access to stored data, whereas in a NAS configuration, the storage server provides clients with file-level access to stored data. Thus, the techniques introduced above are not limited to use in a file server or in a NAS environment.

[0048] This invention has been described with reference to specific exemplary embodiments thereof. It will, however, be evident to persons having the benefit of this disclosure that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the invention. The specification

and drawings are accordingly to be regarded in an illustrative, rather than in a restrictive sense.